

# PATENT ABSTRACTS OF JAPAN

(11)Publication number : 04-158397

(43)Date of publication of application : 01.06.1992

(51)Int.Cl.

G10L 3/00

(21)Application number : 02-284965

(71)Applicant : A T R JIDO HONYAKU DENWA  
KENKYUSHO:KK

(22)Date of filing : 22.10.1990

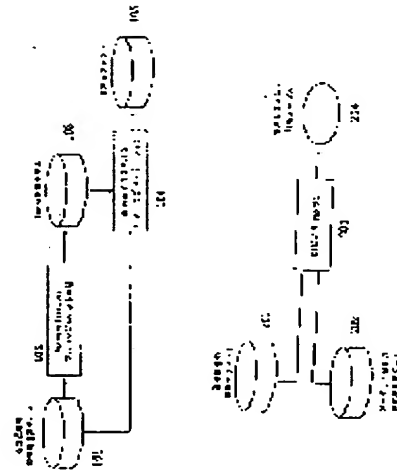
(72)Inventor : ABE MASANOBU  
SAGAYAMA SHIGEKI

## (54) VOICE QUALITY CONVERTING SYSTEM

### (57)Abstract:

**PURPOSE:** To perform the detailed voice quality conversation by correlating parameters between a reference speaker and a target speaker for each voice segment, performing voice conversation based on the correlation, and efficiently expressing the spectrum of the voice.

**CONSTITUTION:** The voice data 101 generated by a reference speaker are A/D-converted and LPC-analyzed. The learning by the forward/backward algorithm is performed with the data, and the HMM voice model 103 for each voice is obtained. Recognition is performed by a segmentation processing section 104 via the Viterbi algorithm with the model 103 to obtain a voice segment 105. Voice segment correlating processing is performed by the processing section 104 with the voice segment 105. The voice segment 201 of the reference speaker and the voice with the same content generated by the target speaker are sent to a DP correlation processing section 203 as learning voice data 202, they are correlated for each frame, and a voice segment correlation table 204 is obtained.



BEST AVAILABLE COPY

## LEGAL STATUS

[Date of request for examination]

[Date of sending the examiner's decision of rejection]

[Kind of final disposal of application other than the examiner's decision of rejection or application converted registration]

[Date of final disposal for application]

[Patent number]

[Date of registration]

[Number of appeal against examiner's decision]

**THIS PAGE BLANK (USPTO)**

of rejection]

[Date of requesting appeal against examiner's  
decision of rejection]

[Date of extinction of right]

Copyright © 2006 by NCIP

**THIS PAGE BLANK (USPTO)**

## ⑫ 公開特許公報(A) 平4-158397

⑤ Int. Cl.<sup>5</sup>

識別記号

庁内整理番号

⑬ 公開 平成4年(1992)6月1日

G 10 L 3/00

E

8622-5H

審査請求 有 請求項の数 3 (全4頁)

## ⑭ 発明の名称 声質変換方式

⑮ 特 願 平2-284965

⑯ 出 願 平2(1990)10月22日

特許法第30条第1項適用 平成2年9月19日、社団法人日本音響学会発行の「日本音響学会 平成2年度秋季研究発表会講演論文集Ⅰ」に発表

⑰ 発 明 者 阿 部 匡 伸 京都府相楽郡精華町大字乾谷小字三平谷5番地 株式会社  
エイ・テイ・アール自動翻訳電話研究所内

⑱ 発 明 者 嵯 峨 山 茂 樹 京都府相楽郡精華町大字乾谷小字三平谷5番地 株式会社  
エイ・テイ・アール自動翻訳電話研究所内

⑲ 出 願 人 株式会社エイ・テイ・アール自動翻訳電話研  
究所 京都府相楽郡精華町大字乾谷小字三平谷5番地

⑳ 代 理 人 弁理士 深見 久郎 外2名

## 明 細 書

## 1. 発明の名称

声質変換方式

## 2. 特許請求の範囲

(1) デジタル化された音声に対しデジタル信号処理を行なってパラメータを抽出し、そのパラメータを制御して音声の声質変換を行なう声質変換方式において、

基準の話者とターゲットとなる話者の間で音声セグメントを単位としてパラメータの対応づけを行ない、この対応づけに基づいて声質変換を行なうことを特徴とする声質変換方式。

(2) 前記話者とターゲット話者間の音声セグメントの対応は、一定の音声データを用いた学習により求め、これに基づいて声質の変換を行なうことを特徴とする、請求項第1項記載の声質変換方式。

(3) さらに、学習の際にDPマッチングの対応づけにより基準話者とターゲット話者の音声セグメントの対応づけを求め、声質の変換を行な

うことを特徴とする、請求項第2項記載の声質変換方式。

## 3. 発明の詳細な説明

[産業上の利用分野]

この発明は声質変換方式に関し、特に、音声セグメントを単位とし、音声の音質を特定の話者の声質に似せたり、規則合成システムから多種類の音質の音声を出力するような声質変換方式に関する。

[従来の技術および発明が解決しようとする課題]

従来より、音声の音質を特定の話者の声質に似せたり、規則合成システムから多種類の音質の音声を出力するために声質変換方式が用いられている。この場合、音声のスペクトルに含まれる個性は、ごく一部のパラメータ(たとえばスペクトルパラメータの中のフォルマント周波数やスペクトル全体の傾きなど)を制御し、声質を変換していた。しかしながら、これらの従来の方式では、たとえば男女声変換のような大雑把な声変換しかすることができない。また、大雑把な声質変換を

行なうにしても、声質を特徴づけるパラメータの変換規則の求め方が確立されておらず、ヒューリスティックな手順を必要とするという問題点があった。

それゆえに、この発明の主たる目的は、音声セグメントを用いて個人のスペクトル空間を表現し、この空間の対応づけにより声質の変換を行なうことによって、詳細な声質変換を可能にし得る声質変換方式を提供することである。

#### [課題を解決するための手段]

この発明はデジタル化された音声に対しデジタル信号処理を行なってパラメータを抽出し、このパラメータを制御して音声の声質変換を行なう声質変換方式であって、基準の話者とターゲットとなる話者の間で音声セグメントを単位としてパラメータの対応づけを行ない、この対応づけに基づいて声質変換を行なうように構成される。

より好ましくは、基準話者とターゲット話者間の音声セグメントの対応を一定の音声データを用いた学習により求め、これに基づいて声質の変換

を行なう。より好ましくは、学習の際にDPマッチングの対応づけにより基準話者とターゲット話者の音声セグメントの対応づけを求め、声質の変換を行なう。

#### [作用]

この発明に係る声質変換方式は、基準話者とターゲットとなる話者の間で音声セグメントを単位としてパラメータの対応づけを行ない、この対応づけに基づいて声質変換を行なうことにより、音声のスペクトルを効率よく表現できる。特に、音声セグメントは音声全体を離散的に表現する1つの手法であり、動的な特徴も含まれているので、従来のようにスペクトルの情報の一部のみを制御する場合に比べて詳細な声質変換が可能となる。

#### [発明の実施例]

第1図はこの発明の一実施例におけるセグメンテーション処理部の概略ブロック図であり、第2図は音声セグメント対応づけ処理のブロック図であり、第3図は声質変換合成のためのブロック図である。

この発明の一実施例では、音声セグメントとして音素を採用し、セグメンテーション処理と音声セグメントの対応づけと声質変換合成との3つのステップからなる。第1図に示すセグメンテーション処理部では、学習用音声を音声セグメントに分割するための処理が音声セグメンテーション処理部によって行なわれる。第1図に示した音声セグメンテーション処理部は、隠れマルコフモデル(HMM)を用いた例である。基準話者(この話者の音声に変換される)が発声した音声データ101はA/D変換された後に、LPC分析される。このデータを用いてforward-Backwardアルゴリズム(L. E. Baum, "An inequality and associated maximization technique in statistical estimation for probabilistic function of Markov process, "Inequalities, 3, pp. 1-8, 1972.)による学

習102が行なわれ、音素ごとのHMM音素モデル103が得られる。HMM音素モデル103を用いてViterbiアルゴリズムによるセグメンテーション処理部104によって認識が行なわれ、音声セグメント105が得られる。なお、Viterbiアルゴリズムについては、中川、「確立モデルによる音声認識」、電子情報通信学会編、pp. 44-46, 1988.に記載されている。

上述のごとくして求められた音声セグメントを用いて、第2図に示した音声セグメント対応付け処理部によって音声セグメント対応づけ処理が行なわれる。すなわち、基準話者の音声セグメント201と、ターゲット話者(その人が発声したように変換したい話者)が発声した同一内容の発声が学習用音声データ202とされ、DPによる対応づけ処理部203に与えられる。なお、基準話者の音声は、第1図に示したセグメンテーション処理部でセグメント化されているものとする。ターゲット話者の音声セグメントは次のようにして

求められる。

まず、両話者の発生した音声データの間で、D Pによる対応づけ処理部203によってフレームごとの対応づけが求められる。D Pによる対応づけ処理部203については、迫江、千葉、「動的計画法を利用した時間正規化に基づく連続音声認識」、音響誌、27、9、pp. 483-490 1971に記載されている。

次に、この対応づけに従って、基準話者の音声セグメント境界がターゲット話者の音声のどのフレームに対応しているかが調べられ、対応したフレームがターゲット話者の音声セグメント境界として定められる。このようにして、音声セグメント対応テーブル204が得られる。

次に、第3図に示した声質変換合成処理部によって声質変換合成が行なわれる。基準話者の音声データは音声分析処理部301に与えられてL P C分析された後、第1図に示したセグメンテーション処理部で作成された基準話者のHMM音楽モデル302を用いて、V i t e r b iアルゴリズム

ム音楽モデルを用いて、セグメンテーション処理部303によってV t e r b iアルゴリズムによるセグメンテーションが行なわれる。次に、このセグメンテーションされた音声に最も近い音声セグメントが最適音声セグメントの探索処理部305によって基準話者の学習用音声セグメント304の中から選択される。選ばれた基準話者の音声セグメントに対応する音声セグメントは、ターゲット話者の学習用音声セグメント308から、第2図に示した音声セグメント対応づけ処理部で作成された音声セグメント対応テーブル306を用いて、音声セグメントの入れ換処理部307によって求められる。最後に、音声合成処理部309によって、求められた音声セグメントを用いて合成され、変換された音声が出力される。

#### [発明の効果]

以上のように、この発明によれば、基準の話者とターゲットとなる話者の間で音声セグメントを単位としてパラメータの対応づけを行ない、この対応づけに基づいて声質変換を行なうことができ

る。特に、音声セグメントは音声全体を離散的に表現する1つの手法であり、音声符号化、規則合成の研究で裏付けられているように、音声のスペクトルを効率よく表現でき、スペクトルの情報の一部のみを制御する従来例に比べて、詳細な声質変換が可能となる。しかも、音声セグメント内には音声の静的な特徴ばかりでなく、動的な特徴も含まれているので、音声セグメントを単位として用いることにより、動的な特徴が変換可能となり、より詳細な個人性の表現が可能となる。さらに、この発明によれば、学習用データさえあれば声質変換することが可能であるため、不特定多数の音声の個人性を得ることが容易となる。

#### 4. 図面の簡単な説明

第1図はこの発明の一実施例における音声セグメンテーション処理部の概略ブロック図である。第2図は音声セグメント対応付け処理部のブロック図である。第3図は声質変換合成のためのブロック図である。

図において、101は基準話者の学習用音声デ

ータ、102は学習処理部、103はHMM音楽モデル、104はセグメンテーション処理部、105は音声セグメント、201は基準話者の音声セグメント、202はターゲット話者の学習用音声データ、203は対応づけ処理部、204は音声セグメント対応テーブル、301は音声分析処理部、302は基準話者のHMM音楽モデル、303はセグメンテーション処理部、304は基準話者の音声セグメント、305は探索処理部、306はセグメント対応テーブル、307は入れ換処理部、308はターゲット話者の音声セグメント、309は音声合成処理部を示す。

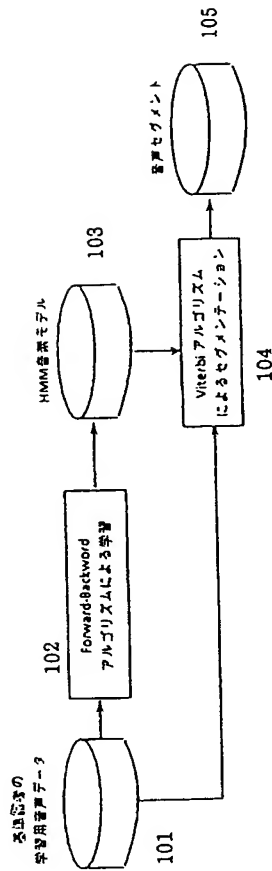
特許出願人 株式会社エイ・ティ・アール自動  
翻訳電話研究所

代理人 弁理士 深見 久郎

(ほか2名)

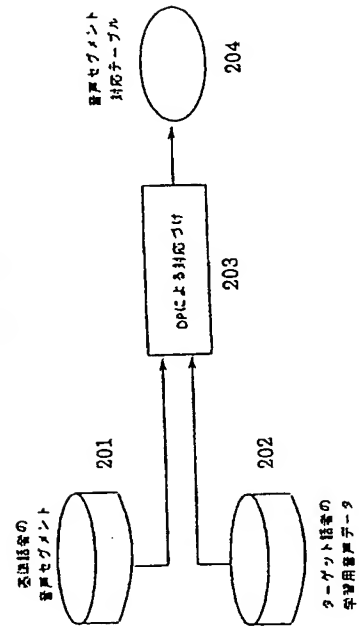


第 1 図



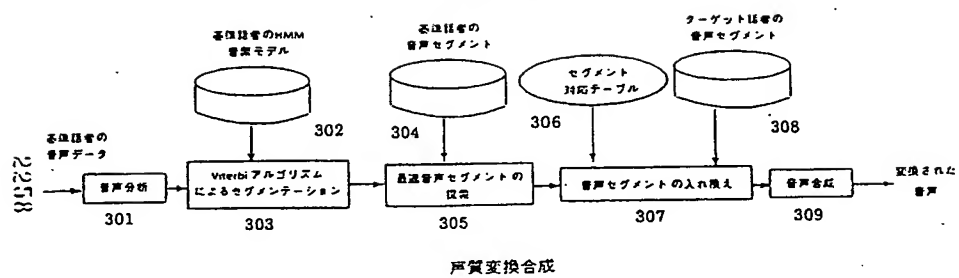
音声セグメンテーション処理部

第 2 図



音声セグメント対応づけ処理部

第 3 図



声質変換合成